



Visibilité de la production scientifique : une étude scientométrique au Pôle Universitaire Lyonnais

Camille Prime-Claverie, Stéphanie Pouchot, Thierry Lafouge, Benoît Epron

► To cite this version:

Camille Prime-Claverie, Stéphanie Pouchot, Thierry Lafouge, Benoît Epron. Visibilité de la production scientifique : une étude scientométrique au Pôle Universitaire Lyonnais. "Partage de l'information dans un monde fragmenté : Franchir les frontières", 39ème Congrès annuel de l'Association Canadienne des Sciences de l'Information,, May 2007, Université McGill, Montréal (Canada). sic_00150563

HAL Id: sic_00150563

https://archivesic.ccsd.cnrs.fr/sic_00150563

Submitted on 30 May 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Camille Prime-Claverie
Université Paris X – Nanterre, France ; CRIS-SERIES.
F-92001 Nanterre cedex, France

Stéphanie Pouchot, Thierry Lafouge
Université de Lyon, France ; Université Lyon 1, ELICO, EA 4147.
F-69622 Villeurbanne cedex, France

Benoît Epron
Université de Lyon, France ; Enssib, ELICO, EA 4147.
F-69623 Villeurbanne cedex, France

Visibilité de la production scientifique : une étude scientométrique au Pôle Universitaire Lyonnais

Résumé : L'infométrie joue un rôle important dans l'évaluation des organismes de recherche et de leur politique. Dans ce contexte, nous présentons les résultats d'une étude scientométrique sur la production des chercheurs des établissements scientifiques du Pôle Universitaire Lyonnais menée entre 2000 et 2005.

Abstract: Informetrics plays an important part in the evaluation of research organizations as well as their policies. In this context, we will share results from a scientometric study on outputs from University of Lyon scientific research centers between 2000 and 2005.

1. Introduction

L'évaluation des universités en fonction de critères infométriques est aujourd'hui une réalité. Cette approche soulève un certain nombre de problèmes méthodologiques liés au mode de collecte des données, au choix des indicateurs utilisés, aux disciplines elles-mêmes ainsi qu'aux politiques de recherche nationales. Un indicateur ne prend en effet sens que dans un contexte donné. Néanmoins, à l'heure de la mondialisation et devant la mobilité croissante des chercheurs, les études comme celle de Shanghai¹, amènent la communauté scientifique à s'interroger sur la visibilité de la recherche.

Dans ce contexte, nous avons mené une étude scientométrique sur la production des chercheurs des établissements scientifiques du Pôle Universitaire Lyonnais² (PUL) entre 2000 et 2005.

Dans la première partie de cet article, nous revenons sur le rôle de l'infométrie dans l'évaluation des politiques de recherche et des structures de recherche universitaires. Nous expliquons également les réactions suscitées par le classement de Shanghai puis nous donnons un exemple des difficultés à évaluer la recherche en France.

Dans un second temps, nous présentons l'étude que nous avons menée sur la visibilité des établissements scientifiques du PUL : nous détaillons les sources utilisées, la constitution de notre corpus ainsi que les résultats.

¹ <http://ed.sjtu.edu.cn/ranking.htm> (page visitée le 05/04/07)

² <http://www.universite-lyon.fr/> (page visitée le 05/04/07)

Pour conclure, nous replaçons les enjeux de ce travail dans leur contexte.

2. Scientométrie et évaluation : la polémique suscitée par le classement de Shanghai

La scientométrie est l'application de méthodes statistiques à des données quantitatives (bibliographiques, humaines, économiques, *etc.*) dans le but de mesurer et caractériser l'état de la science. Il est habituel de situer les débuts de la scientométrie au début des années 1960 avec les travaux théoriques de Price (Price, 1963). En 1963, Garfield crée le Science Citation Index³ et fournit ainsi à la scientométrie un instrument de mesure de la Science qui confère à l'article scientifique un rôle central et est utilisé pour l'évaluation des chercheurs (Garfield, 1981). L'auteur intéressé par ces aspects historiques peut se référer par exemple à l'article de Xavier Polanco sur les origines de la scientométrie (Polanco, 1995).

La plupart du temps, les articles de référence dans ce domaine ignorent les travaux de James McKeen Cattell. Psychologue américain du début du XX^{ème} siècle, il a pourtant initié les premières statistiques systématiques sur la production des chercheurs en 1906. Éditeur de la revue *Science* de 1895 à 1944, Cattell a développé le premier répertoire des chercheurs américains actifs et en a dégagé les plus éminents (Godin, 2006).

La discipline est reconnue comme telle suite à la création de la revue *Scientometrics* en 1978. Puis, c'est lors du premier colloque international dans ce domaine en 1987 que le terme « infométrie » va être défini comme étant l'ensemble des activités métriques relatives à l'information, couvrant aussi bien la bibliométrie⁴ que la scientométrie.

Depuis, de nombreuses études évaluant la science (productions scientifiques, disciplines...) ont été menées en soutien à la conduite de politiques de recherche (Ingwersen, 2001 ; Coccia, 2005). Des organismes parapublics chargés de conduire ces études et de produire des indicateurs mondiaux ont ainsi vu le jour comme le CWTS⁵ pour les Pays Bas et la Belgique, le CINDOC⁶ pour l'Espagne et l'OST⁷ en France. Par ailleurs, des pays émergents comme le Brésil (Gusmao, 2006), l'Inde (Ravichandra Rao, 1999) ou la Chine (Jin et al., 2003) sont également devenus des acteurs importants de la scientométrie. L'enjeu pour ces pays est de pouvoir se situer par rapport à la recherche mondiale, de mieux connaître leurs forces régionales et d'investir dans les secteurs en expansion.

Les résultats d'une étude chinoise ont récemment donné lieu à polémique. Publié intégralement sur le site Web de l'Institute of Higher Education de Shanghai, le fameux classement mondial des universités (Liu, 2005) est en effet critiquable à plus d'un titre. Non validée par la communauté scientométrique mais néanmoins publiée, cette étude a

³ Base de données scientométrique de l'ISI

⁴ « Ensemble des méthodes et techniques quantitatives – de type mathématiques / statistiques – susceptibles d'aider à la gestion des bibliothèques et d'une manière très générale des divers organismes ayant à traiter de l'information » (Noyer, 1995).

⁵ Centre for Science and Technology Studies, <http://www.cwts.nl/scripts/index.pl> (page visitée le 05/04/07)

⁶ Centro Informacion y Documentacion Cientifica, <http://www.cindoc.csic.es/eng/principal.html> (page visitée le 05/04/07)

⁷ Observatoire des Sciences et Techniques, <http://www.obs-ost.fr/> (page visitée le 05/04/07)

eu un fort impact, en particulier en France, en raison du très mauvais classement des universités françaises.

Des scientomètres réputés, comme Van Raan et Zitt, ont vivement réagi à cette étude (Van Raan, 2006 ; Zitt et Filliatreau, 2005). Nous revenons ici rapidement sur ces critiques d'ordre méthodologique. Tout d'abord, les objets comparés (les universités du monde entier quelle que soit leur taille) sont-ils vraiment comparables ? D'autre part, les techniques à mettre en œuvre et les précautions à prendre lorsque l'on utilise les méthodes de citations pour évaluer la recherche (Moed, 2005) ne sont-elles pas ignorées ici ? En outre, les indicateurs présentés sont-ils utilisés à bon escient ? Il ne semble pas surprenant de trouver en tête du classement les quatre plus grandes universités états-uniennes puisque les indicateurs ont un « effet de taille » : la grande majorité des indicateurs dépend de la taille de l'acteur observé et la reflète, particulièrement lorsqu'ils sont corrélés entre eux.

Par ailleurs, compte tenu de la variété des appellations d'affiliations pour une même entité de recherche, une question mérite d'être posée : les résultats sont-ils réellement représentatifs de la production des chercheurs ? Une étude plus fine sur la production scientifique médicale dans une grande université française menée par une équipe de notre laboratoire (Bador et Lafouge, 2005) a en effet mis en évidence l'inconstance des chercheurs dans la mention de leur affiliation lorsqu'ils publient un article. Sur un échantillon de trois cents notices bibliographiques, quarante-quatre types de libellés d'adresses différents pour un même établissement ont été repérés. Cet aspect n'est pas lié aux bases de l'ISI puisque ces travaux portaient sur un échantillon extrait des bases médicales Medline, Embase et Biomed. Ce manque de visibilité est bien réel et est dû en partie à l'organisation de la recherche et plus particulièrement de la recherche médicale comme l'avait souligné Van Raan précédemment (Van Raan, 2006).

Ces dernières années, la communauté scientifique française a pris conscience de la place cruciale de l'évaluation. En octobre 2006, l'OST a lancé un chantier d'uniformisation des adresses des organismes de recherche français⁸. Ce projet est commandité par le Ministère de l'Education Nationale de l'Enseignement Supérieur et de la Recherche, en concertation avec l'éditeur du WOS⁹. L'Université Claude Bernard Lyon 1 s'en préoccupe quant à elle depuis 2000, sous l'impulsion d'un groupe de travail interne, EPSILON, dont nous présentons ci-après une partie des résultats de recherche.

3. Étude scientométrique des établissements scientifiques du PUL

Le PUL regroupe seize établissements lyonnais d'enseignement supérieur et de recherche. En émanant chaque année plusieurs centaines de publications, dans des domaines aussi variés que l'archéologie, la psychologie ou la physique nucléaire. Compte tenu de la diversité des disciplines concernées, il est à l'heure actuelle extrêmement difficile d'identifier et recenser de manière précise la production des chercheurs du PUL. Néanmoins, comme nous venons de le voir, les enjeux sont importants. Il s'agit d'une part de comprendre les activités scientifiques et technologiques des établissements en vue de prises de décisions politiques et budgétaires. D'autre part, cela permet de connaître le rayonnement du pôle universitaire lyonnais tel qu'il est perçu par la communauté internationale.

⁸ <http://www.obs-ost.fr/normadresses.phtml> (page visitée le 05/04/07)

⁹ Web Of Science, <http://scientific.thomson.com/products/wos/> (page visitée le 05/04/07)

Dans cette partie, nous revenons sur le terrain de notre étude. Nous détaillons ensuite notre méthodologie de constitution de corpus, d'organisation et d'analyse de données avant d'exposer nos résultats.

3.1. Contexte de l'étude

Faisant partie d'une université de Sciences et Médecine et ayant une bonne connaissance du terrain, nous avons limité notre étude à la production des cinq établissements de sciences dures et de médecine du PUL :

- Université Claude Bernard Lyon 1,
- Institut National des Sciences Appliquées,
- École Centrale de Lyon,
- École Normale Supérieure de Sciences,
- École Nationale Vétérinaire de Lyon.

Par ailleurs, le nombre important d'unités de recherche¹⁰ de ces établissements rend difficile, sinon impossible, une étude basée sur les rapports propres de ces unités. Nous avons donc choisi d'évaluer leur visibilité à travers les bases de données internationales de l'ISI. Depuis longtemps, ces bases permettent en effet des analyses scientométriques car l'affiliation précise de chaque auteur y est disponible.

Nous présentons les résultats construits sur des données couvrant la période de 2000 à 2005.

La base de données bibliographique utilisée pour cette étude est la base des *Current Contents* produite par l'ISI. Il s'agit d'une base de données indexant environ 8000 revues de rang A (revues internationales à comité de lecture). Les publications sont classées dans sept sections thématiques nommées CC (pour « Classification Codes ») :

- Agriculture, Biology and Environmental Sciences,
- Arts and Humanities,
- Clinical Medicine,
- Engineering, Computing and Technology,
- Life Sciences,
- Physical, Chemical and Earth Sciences,
- Social and Behavioral Sciences.

Cette base ne propose malheureusement aucune indication sur les citations reçues ou émises par les articles ce qui compromet la production d'indicateurs d'impact.

Pour notre étude, nous avons adopté une approche documentaire classique, à savoir la constitution d'un corpus raffiné et filtré afin d'en extraire des résultats statistiques.

3.2. Constitution du corpus

3.2.1. Bruit et silence

La notion de bruit, c'est-à-dire les articles non pertinents retrouvés suite à une requête, est une propriété inhérente aux systèmes de recherche d'information. Dans notre cas, le bruit représente les articles retrouvés mais non publiés par l'un des établissements du PUL. Dans le cadre d'une recherche documentaire classique, le bruit est généralement engendré

¹⁰ L'université Claude Bernard Lyon 1 comporte à elle seule 102 unités de recherches dont certaines mixtes avec d'autres établissements.

par des problèmes intrinsèques à la langue naturelle, comme l'homonymie ou la polysémie. Dans cette étude, le phénomène de bruit dû aux ambiguïtés linguistiques reste minoritaire. Sa cause principale provient d'un manque de précision au niveau du mode d'interrogation des *Current Contents*. En effet, l'interrogation sur l'affiliation des auteurs se fait en réalité sur un index regroupant plusieurs champs¹¹ dont certains sont multivalués comme le champ *Research Institution* (IN). Il est alors impossible de formuler une requête sur les institutions prises séparément. Ainsi, une requête sur cet index comportant les termes « univ » et « Lyon » retrouve par exemple des co-publications du Cemagref de Lyon avec l'Université de Montpellier.

Le silence concerne les articles publiés par un établissement du PUL et non retrouvés suite à nos requêtes. Réduire ce phénomène est l'une de nos priorités puisqu'il nuit à la bonne visibilité des établissements du PUL. Outre les erreurs habituellement commises par les utilisateurs lors de l'interrogation de la base, le silence est imputable aux erreurs et imprécisions commises aux différents niveaux de la chaîne éditoriale (auteurs, éditeurs scientifiques ou commerciaux, indexeurs). Il provient néanmoins majoritairement d'un manque de précision de la part des auteurs. En effet, les dénominations de leurs établissements et laboratoires d'origine sont rarement normalisées et différent suivant les publications. Ce phénomène est particulièrement important pour les établissements localisés sur plusieurs sites, comme le montrent (Bador et Lafouge, 2005).

3.2.2. Stratégies d'interrogation : formation du corpus

Pour une évaluation précise de la visibilité des établissements du PUL, réduire le bruit et le silence est un objectif primordial. Le repérage du périmètre des publications a été effectué grâce à deux stratégies :

- La première est une recherche sur les affiliations. Cette stratégie offre l'avantage de récupérer assez précisément (avec les limites évoquées ci-dessus) les publications des établissements du PUL. Cependant, il est impossible de prévoir l'ensemble des affiliations sous lesquelles publient les auteurs des établissements étudiés, et ceci peut être une cause de silence.
- La seconde est une recherche sur l'origine géographique des publications. Elle nous apporte un plus grand nombre de réponses qui devront être filtrées pour valider leur appartenance à l'un des cinq établissements étudiés.

Les équations de recherche utilisées pour chacune des deux stratégies sont présentées dans le tableau ci-dessous.

¹¹ L'index interrogeable AD regroupe les champs suivants : CA (Corporate Author), EM (Email Address), IN (Research Institution), RP (Reprint Author), AD (Address Information)

Équation 1 : stratégie de recherche sur les affiliations	Équation 2 : stratégie de recherche sur l'origine géographique
<pre>(((insa or (inst* nat* sci* appl*)) and (lyon or villeurbanne)) or (ucbl or ucb or (univ* claude bernard) or doua or univ* or (obs* lyon)) and (lyon or villeurbanne or bron or (pierre benite) or (bourg en bresse) or (genis laval))) or ((((ctr hosp) or chu or hop*) and (lyon or oullins)) or (lyon sud)) or (((fac and (med or pharm)) or (grange blanche) or (rockefeller) or (laennec) or (med sch)) and (lyon or villeurbanne or oullins))) or (((ecole cent*) or ecl) and (lyon or ecully)) or (((ecole norm* super*) or ens) and lyon) or ((ecole vet*) and lyon)) in ad</pre>	<pre>((Lyon or Villeurbanne or (Pierre Benite) or Ecully or Oullins or (Genis laval) or (Marcy and Etoile) or (Bourg en Bresse) or Bron or (Seyne and Tamaris)) and France) in ad</pre>
12 888 références datées de 2000 à 2003 retrouvées (interrogation réalisée le 2 février 2004).	14 966 références datées de 2000 à 2003 retrouvées (interrogation réalisée le 8 mars 2004).

Tableau 1 : Equations de recherche

Les références retrouvées sont rapatriées au format complet et constituent nos deux sous-corpus de travail.

3.3. Organisation du corpus : création des bases de données

Pour chacun des deux sous-corpus obtenus, nous avons vérifié l'appartenance des articles à l'un des cinq établissements du PUL par deux filtrages successifs de plus en plus précis. Ceci permet de repérer le bruit et de réduire le silence. Ce filtrage a nécessité la création d'une base de données. En effet, comme nous l'avons signalé, le champ *Research Institution* (IN) de la base des *Current Contents* (champ reprenant les affiliations de tous les auteurs) étant multivalué, il peut prendre une ou plusieurs valeurs. Il est alors nécessaire d'éclater ce champ afin d'isoler chaque institution pour le filtrage.

Les deux sous-corpus sont organisés au sein de deux bases de données qui présentent la même structure (Figure 1). Ces bases comportent cinq tables : les trois premières représentent respectivement les trois entités suivantes : référence, institution, établissement. Les deux autres donnent les relations entre ces entités.

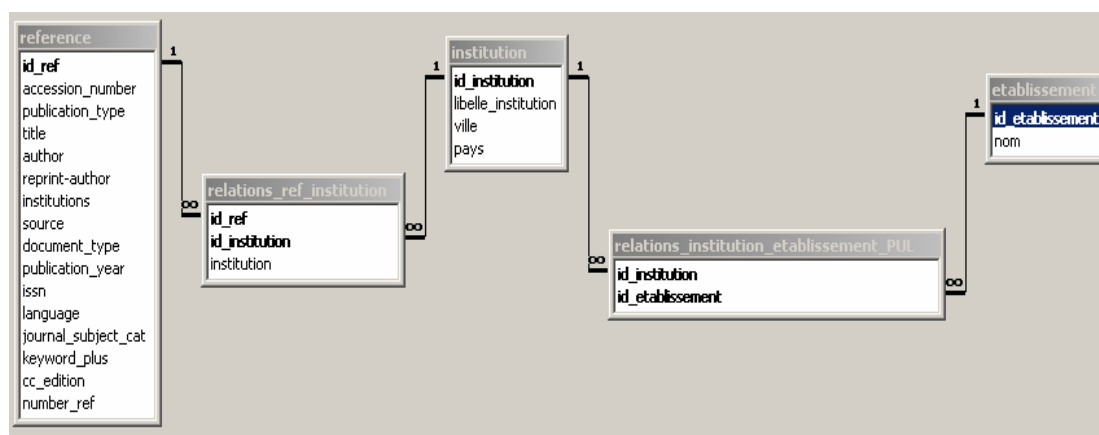


Figure 1 : Structure de la base de données

Toute la difficulté de ce travail réside dans la construction de la table relationnelle entre les institutions localisées en région lyonnaise et les établissements du PUL étudiés.

L'affectation d'une institution à un ou plusieurs établissements du PUL se fait grâce à une bonne connaissance du terrain. Nous avons mis au point deux filtres qui analysent automatiquement les termes du libellé des institutions.

- Le premier filtre concerne les différentes appellations des établissements (noms complets, noms abrégés, sigles, etc.). Il examine le libellé de chaque institution et vérifie s'il contient ou non les appellations utilisées pour nommer les établissements étudiés.
- Le second filtre étudie les institutions non affectées par le premier filtre. Il vérifie si celles-ci sont des laboratoires mixtes du CNRS¹². Le cas échéant, les institutions sont affectées à leur(s) organisme(s) de tutelle lorsque celui-ci est l'un des établissements étudiés.

L'ensemble du processus de recherche et de filtrage est présenté dans la *Figure 2*.

L'objectif du filtrage est d'identifier le bruit (e) parmi les articles retrouvés par une stratégie (c). Il s'agit que l'ensemble des articles non reconnus par f_2 soit le plus proche possible de (e). La différence entre (c_{f_2}) et (e) s'apparente également à une forme de silence. La réduction du silence entre les deux filtrages est matérialisée par l'ensemble (c_{f_2}).

¹² Centre National de la Recherche Scientifique, <http://www.cnrs.fr/> (page visitée le 05/04/07)

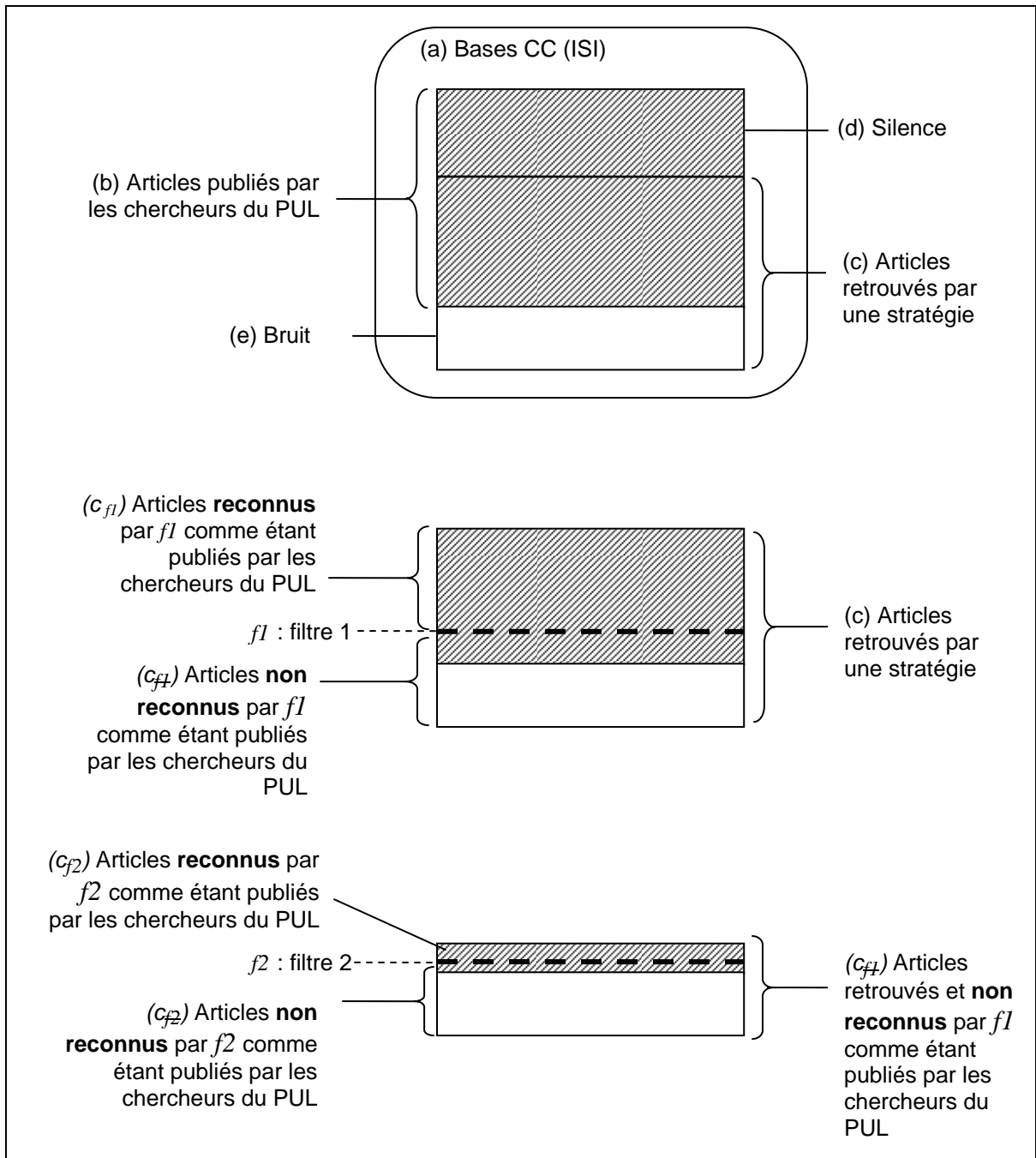


Figure 2 : Processus de filtrage

Citons quatre exemples de libellés contenus dans des références bibliographiques analysées par les filtres :

- (a) Ecole Cent Lyon, Dept Math & Informat, CNRS, UMR 5585, MAPLY
- (b) Inst Natl Sci Appl, Ctr Math, MAPLY
- (c) Univ Lyon 1, CNRS, MAPLY
- (d) MAPLY, UMR 5585

À l'issue du premier filtre les trois premières institutions sont respectivement affectées aux établissements *Ecole Centrale de Lyon*, *INSA*, *Université Claude Bernard Lyon 1* et ne seront pas analysées par le second filtre. La quatrième institution est analysée par le second filtre et est affectée, selon la méthode du *compte de présence* décrite au § 3.4, aux

trois organismes de tutelle du laboratoire MAPLY (laboratoire de Mathématiques appliquées) qui sont : l'Ecole Centrale de Lyon, l'INSA et l'Université Claude Bernard Lyon 1.

Le premier filtrage utilise entre deux et quatre expressions régulières¹³ pour l'identification des établissements suivants : INSA, Ecole Centrale de Lyon, Ecole Normale Supérieure, Ecole Nationale Vétérinaire de Lyon. En ce qui concerne l'Université Claude Bernard Lyon 1, 30 expressions régulières sont nécessaires étant donné le nombre de variantes possibles pour désigner cette université.

Cela montre qu'un tel filtrage peut difficilement être entrepris dans le cadre d'une évaluation à grande échelle de plusieurs établissements, *a fortiori* à l'échelle internationale.

3.4. Comparaison des résultats issus des deux stratégies de recherche entre 2000 et 2003

Conformément à la méthodologie utilisée par l'OST, le principe de comptage pour les établissements est celui du *compte de présence*. Sur ce principe sont comptabilisées les publications où l'établissement apparaît au moins une fois. Il n'est pas fait de différence entre la présence d'un seul ou de plusieurs laboratoires de l'établissement.

Le tableau ci-dessous indique le nombre de publications retrouvées pour l'ensemble des établissements du PUL, et pour chaque stratégie (équation 1 et équation 2) après l'application de chaque filtre.

		2000 à 2003	2000	2001	2002	2003 ¹⁴
Équation 1	filtre 1	9 797	2 538	2 458	2 501	2 300
	filtres 1 et 2	10 004	2 590	2 511	2 549	2 354
Équation 2	filtre 1	10 151	2 581	2 506	2 541	2 523
	filtres 1 et 2	10 585	2 686	2 605	2 651	2 643

Tableau 2 : Nombre de publications entre 2000 et 2003 avec différentes stratégies

3.4.1. Évaluation du bruit

Le Tableau 2 met en évidence la réduction du bruit réalisée grâce au filtrage. Le filtrage du corpus obtenu par la seconde équation portant sur les origines géographiques, montre la présence d'un bruit important dans le corpus : sur les 14 966 références retrouvées (cf. Tableau 1), seules 10 585 appartiennent aux établissements du PUL selon nos critères, soit une réduction de 29,2 % du bruit. Le filtrage montre également la présence d'un bruit important dans le premier corpus qui, quant à lui, a été construit par une équation très précise sur les affiliations. En effet, sur les 12 888 références retrouvées, seules 10 004 proviennent des établissements du PUL, soit 22 % des notices n'appartenant pas au PUL.

3.4.2. Évaluation de la réduction du silence

Les résultats montrent que l'application du second filtre et la recherche par la seconde stratégie permettent de retrouver plus d'articles et d'atténuer le silence.

- Réduction du silence après l'application du second filtre :

1. équation 1 : $1 - (9797/10004) = 2 \%$

¹³ Les expressions régulières permettent de retrouver des chaînes de caractères dans un texte. Dans notre cas, elles sont utilisées dans des scripts PERL.

¹⁴ Année incomplète : les chiffres donnés pour l'année 2003 sont à examiner avec précaution. En effet, il existe un décalage entre l'apparition d'une publication et son enregistrement dans la base. On considère qu'il faut attendre 30 semaines d'une année $n+1$ pour avoir la totalité des publications de l'année n dans la base.

2. *équation 2* : $1-(10151/10585) = 4,1 \%$
- Réduction du silence pour la seconde stratégie
 1. *filtre 1* : $1-(9797/10151) = 3,5 \%$
 2. *filtre 2* : $1-(10004/10585) = 5,5 \%$
- Réduction maximum du silence entre
 - le nombre de publications retrouvées par l'équation 1 avec l'application du filtre 1 (9 797)
 - et le nombre de publications retrouvées par l'équation 2 avec l'application du filtre 1 et 2 (10 585)
 => Cette réduction est de : $1-(9797/10585) = 7,4 \%$

Ces résultats montrent que les corpus obtenus par les deux stratégies sont très bruités et nécessitent un filtrage des notices. De plus, la seconde stratégie apporte 581 notices supplémentaires, soit une réduction de 5,5 % du silence entre les deux stratégies. C'est pourquoi par la suite, la stratégie d'interrogation retenue pour le repérage des publications sera la seconde.

3.5. Résultats pour la période 2000-2005

La stratégie d'interrogation étant établie, nous sommes en mesure de proposer des statistiques plus fines pour la période de 2000 à 2005. Nous avons bien sûr la possibilité de fournir les mêmes données pour chaque établissement et pour chaque année. Pour des raisons de confidentialité des données, nous restons ici à un niveau macroscopique.

3.5.1. Les publications des établissements scientifiques du PUL entre 2000 et 2005

Le tableau ci-dessous indique le nombre de publications récoltées entre 2000 et 2005 pour l'ensemble des établissements du PUL. Nous avons appliqué ici la stratégie de recherche décrite précédemment :

Années	2000	2001	2002	2003	2004	2005	Total 2000-2005
Équation 2 Filtres 1 et 2	2 686	2 605	2 651	2 806	2 943	3 052	16 743

Tableau 3 : Nombre de publications du PUL entre 2000 et 2005

Une limite à ces résultats réside dans le recours exclusif à la base *Current Contents*. Or, toutes les disciplines ne sont pas représentées de manière équivalente dans cette base. Par ailleurs, cette base indexe principalement les articles de revues scientifiques de rang A. Les autres formes de publication (actes de conférences, brevets, ouvrages...) ne sont pas prises en compte.

Les résultats ci-dessous traduisent donc la lisibilité du PUL dans cette base de l'ISI.

3.5.2. Résultats par disciplines de l'ISI

Le principe de comptage utilisé pour les disciplines est fractionnaire : une publication qui est assignée à n disciplines compte pour $1/n$ dans chaque discipline.

Le tableau ci-dessous présente pour la période 2000-2005 le nombre de publications par CC (Classification Codes).

Classification Code	2000	2001	2002	2003	2004	2005	Total	%
Agriculture-Biology-and-Environmental-Sciences	184,00	179,83	179,33	195,33	190,33	201,83	1 130,67	6,75%
Arts-and-Humanities	6,00	6,00	3,00	18,50	8,50	6,50	48,50	0,29%
Clinical-Medicine	450,50	472,83	454,83	491,50	546,50	578,83	2 995,00	17,89%
Engineering-Computing-and-Technology	290,00	311,50	332,50	361,50	369,00	387,50	2 052,00	12,26%
Life-Sciences	850,50	773,17	780,67	802,33	819,33	794,67	4 820,67	28,79%
Physical-Chemical-and-Earth-Sciences	884,50	843,33	880,83	912,83	967,33	1049,83	5 538,67	33,08%
Social-and-Behavioral-Sciences	20,50	18,33	19,83	24,00	42,00	32,83	157,50	0,94%
Total	2 686,00	2 605,00	2 651,00	2 806,00	2 943,00	3 052,00	16 743,00	100%

Tableau 4 : Nombre de publications pour 2000-2005 suivant les différentes CC de l'ISI

Ces données caractérisent la production disciplinaire du PUL, en fonction de la classification de l'ISI. Elles sont difficilement lisibles car ces catégories ne correspondent pas totalement à la réalité de la recherche nationale. Par exemple, les *Mathématiques* n'apparaissent pas en tant que telles, elles font ici partie de la catégorie *Physical, Chemical et Earth Sciences* alors qu'en France, elles constituent un champ disciplinaire à part entière (25^{ème} section du Conseil National des Universités).

C'est pourquoi nous avons demandé à l'OST de retraiter nos données de manière à ce qu'elles soient ventilées dans ses catégories propres, plus significatives pour la recherche française.

Les valeurs du *Tableau 5* ne sont pas consolidables sur un ensemble de disciplines. En effet, les publications sont réparties selon un « compte de présence » et peuvent apparaître dans plusieurs catégories. Les totaux affichés à la ligne Toutes disciplines montrent pour chaque année le nombre de références analysées par l'OST. Nous pouvons par ailleurs remarquer que le nombre total de publications pour les années 2000 à 2003 est différent pour les tableaux 4 et 5. En effet, la classification de l'OST basée sur les revues ne permet pas de catégoriser les autres types de documents, comme les ouvrages par exemple. De plus, les valeurs de l'année 2003 ne sont pas complètes du fait de la date de construction des données initiales.

	2000	2001	2002	2003
Biologie fondamentale	573	548	524	537
Recherche médicale	852	809	841	802
Biologie appliquée - écologie	146	141	134	120
Chimie	465	516	487	534
Physique	483	442	455	461
Sciences de l'univers	174	172	180	168
Sciences pour l'ingénieur	257	261	327	297
Mathématiques	91	93	110	117
Multidisciplinaire	23	25	19	17
Toutes disciplines	2 591	2 502	2 560	2 542

Tableau 5 : Volume de publications par discipline (compte de « participation » ou « de présence »)

Le travail de catégorisation des publications est effectué par l'OST selon leur propre plan de classement : Chaque revue est classée dans une ou plusieurs disciplines. A chaque publication est attribué la ou les disciplines de sa source. La catégorie *Multidisciplinaire*, correspond aux revues non spécialisées comme Nature ou Science.

Néanmoins ce travail de retraitement des données par l'OST permet une normalisation des données à l'échelon national. L'intérêt majeur est ici de proposer des outils d'aide à la décision dans le pilotage des universités.

4. Discussion et conclusion

Notre travail dans le cadre du projet EPSILON (recueil de données et filtrage des corpus) ainsi que les indicateurs bibliométriques proposés par l'OST à partir de nos données conduisent à plusieurs pistes de réflexion.

En premier lieu, le décalage entre les résultats issus des bases de l'ISI et leur retraitement par l'OST démontre la nécessité, pour les universités françaises, de recourir à cet observatoire afin d'obtenir des résultats aux formats comparables et donc exploitables pour piloter la recherche. Ainsi, pour être en mesure de proposer ce type d'indicateurs, il est nécessaire d'uniformiser les données à l'échelle nationale en passant par un organisme fédérateur tel que l'OST pour la France.

Ensuite, il convient de s'interroger sur la pertinence des classements internationaux des universités en général. En effet, ils sont en partie construits à partir d'interrogations de bases de l'ISI. Or, nos travaux montrent que l'obtention de résultats significatifs à partir de ces bases nécessite une connaissance fine du terrain et de l'organisation de la recherche au niveau local.

En outre, nos travaux trouvent, d'une manière générale, une résonance forte dans les préoccupations actuelles de l'OST et des autorités de tutelle des universités françaises. En effet, le projet de loi de finances 2007 (PLF) intègre une dimension d'évaluation précise de la production scientifique selon des indicateurs correspondant précisément aux travaux menés dans EPSILON¹⁵. D'une manière plus générale, la mise en place de la Loi Organique relative à la Loi de Finance (LOLF)¹⁶, et donc de ces indicateurs de performance, se déroule dans une logique d'attribution quantitative des moyens.

Les précisions méthodologiques apportées quant à la construction de ces indicateurs nous amènent à plusieurs conclusions.

Dans son projet NORMADRESSES¹⁷, l'OST rejoint clairement la problématique évoquée dans les travaux d'EPSILON, à savoir l'importance de la construction d'une liste précise, vérifiée et à jour des affiliations des chercheurs. L'enjeu est de permettre l'extraction la plus fine et efficace possible des bases bibliographiques.

L'OST préconise en ce sens un travail de terrain fin, au plus près des universités, et a ainsi fourni à chaque organisme de recherche français une liste, à valider, des *Adresses*

¹⁵ http://www.amue.fr/telecharger/publications/PAP_2007_Prog_150.pdf (page visitée le 05/04/07)

¹⁶ <http://www.budget2007.gouv.fr/bleus/BLEUSHTM/DBGPGMOPERATEURPGM150.htm> (page visitée le 05/04/07)

¹⁷ http://www.obs-ost.fr/doc_attach/NoteMethodo1ereVague.PDF (page visitée le 05/04/07)

lui correspondant. Les travaux d'EPSILON confirment cette nécessité d'une connaissance approfondie de la recherche locale.

En outre, il est judicieux, à notre sens, de s'interroger sur la pertinence stratégique d'appuyer les outils d'évaluation utilisés par les institutions pilotes de la recherche sur un outil commercial à dominante anglophone et source unique des données nécessaires à la réalisation des indicateurs utilisés par les pouvoirs publics.

Au-delà de toutes ces problématiques, plane également la question du rôle et de la place de la bibliométrie dans l'évaluation des chercheurs. Pour conclure, afin que cette évaluation ne donne pas lieu à polémique, les règles doivent en être connues et maîtrisées par les différents acteurs du processus (connaissance des sources sur lesquelles sont faites les statistiques, reproductibilité de l'expérience...).

Remerciements

Ce travail a été réalisé avec le soutien financier de la Région Rhône-Alpes à travers deux projets INCA.

Les auteurs tiennent à remercier Jean-François Jal, professeur de physique à l'Université Lyon 1, responsable du projet EPSILON, ainsi que toutes les personnes impliquées dans ce groupe de recherche.

Bibliographie

Bador, Pascal et Lafouge, Thierry. 2005. La difficulté d'accéder aux adresses des chercheurs français dans les banques de données bibliographiques : l'exemple d'une université lyonnaise. *Documentaliste, Sciences de l'information* 43(1): 28-35.

Coccia, Mario. 2005. A New model to evaluate the Scientific Performance of Public Research Institutions. In *Proceedings of 10th International Conference of the International Society for Scientometrics and Informetrics*, Stockholm 2005. Volume 1: 237-241. Karolinska University Press.

Garfield, Eugene. 1981. The 1,000 Contemporary Scientists most-cited 1957-1978 Part I. The basic List and Introduction. *Current Content* (41) : 5-14.

Godin, Benoît. 2006. James Cattell mesure la science. *La Recherche* (401) : 54-57.

Gusmao, Regina. 2006. Production and diffusion of regional ST&I indicators: the state of São Paulo (Brazil) recent experience. In *proceedings First PRIME Indicators Conference*, University of Lugano, Switzerland.

Ingwersen, Peter. 2001. Visibility and Impact of Research in Psychiatry for North European Countries in EU, US and World Context. In *Proceedings of 8th International Conference of the International Society for Scientometrics and Informetrics*, Sydney 2003. Volume 1: 265-274. Published by the bibliometric Research group.

Jin, Bihui, Li, Ling et Rousseau, Ronald. 2003. Production and productivity of Chinese scientists as a function of their age : the period 1995-1999. In *Proceedings of 9th International Conference of the International Society for Scientometrics and*

Informetrics, Pekin 2003: 112-128. Edited by Dalian university of Technology Press Dalian, China.

Liu, Nian Cai et Cheng, Ying. 2005. Academic Ranking of World Universities – Methodologies and Problems. *Higher Education in Europe* 30(2).

Moed, Henk F. 2005. *Citation Analysis in Research Evaluation. Information science and knowledge management*. Springer-Verlag, New York: 350 p.

Noyer, Jean-Max. 1995. Scientométrie, infométrie : pourquoi nous intéressent-elles ? « *Les sciences de l'information : bibliométrie, scientométrie, infométrie* ». *Solaris* (2), Presses Universitaires de Rennes : 175-199.

Polanco, Xavier. 1995. Aux sources de la scientométrie. « *Les sciences de l'information : bibliométrie, scientométrie, infométrie* ». *Solaris* (2), Presses Universitaires de Rennes : 13-76.

Price, Derek John De Solla. 1963. *Little Science, Big Science*. Columbia, New York: 118 p.

Ravichandra Rao, Inna Kedage. 1999. A quantitative Study of Indian Engineering Literature. In *Proceedings of 7th International Conference of the International Society for Scientometrics and Informetrics, Colima, Mexico 1999*: 408-420. Edited by César A. Macias-Chapula.

Van Raan, Anthony. 2006. Fatal attraction : conceptual and methodological problems in the ranking of universities by bibliometric methods, *Scientometrics* 51(1): 335-346.

Zitt, Michel et Filliatreau, Ghislaine. 2005. Big is (made) Beautiful. Somme comments about the Shanghai ranking of world-class universities. In *Proceedings of First International Conference on World Class Universities, Shanghai, June 16-18, 2005*.